Econ 508-A

FINITE DIMENSIONAL OPTIMIZATION - NECESSARY CONDITIONS

Carmen Astorne-Figari
Washington University in St. Louis

August 12, 2010

INTRODUCTION

General form of an optimization problem:

$$\max_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \qquad \text{s.t.} \quad \boldsymbol{x} \in X$$

$\boldsymbol{x} = (x_1, x_2, ..., x_N)$: N-dimensional vector of **choice variables** or **instruments**.

$f(\boldsymbol{x}) = f(x_1, ..., x_N)$: real valued function called **objective function**.

$X \subseteq \mathbb{R}^N$: **constraint set**, **feasible set** or **opportunity set**.

INTRODUCTION

An **optimization problem** in $\mathbb{R}^N$ is one where, given function $f : \mathbb{R}^N \to \mathbb{R}$, we choose the value of $\boldsymbol{x}$ over a given set $X \subseteq \mathbb{R}^N$ that maximizes or minimizes the value of f.

Alternative notation:

$$\max_{\boldsymbol{x} \in X} \quad f(\boldsymbol{x}) \qquad\qquad \min_{\boldsymbol{x} \in X} \quad f(\boldsymbol{x})$$

or

$$\max \{ f(\boldsymbol{x}) \mid \boldsymbol{x} \in X \} \qquad\qquad \min \{ f(\boldsymbol{x}) \mid \boldsymbol{x} \in X \}$$

BASIC DEFINITIONS

Given f and X as above,

The set of attainable values of f on X, or **image of** X **under** f, denoted $f(X)$, is defined by

$$f(X) = \{w \in \mathbb{R} \mid \exists \text{ some } \boldsymbol{x} \in X \text{ such that } f(\boldsymbol{x}) = w\}$$

The **interior of** X is the set defined by

$$\text{int}(X) = \{\boldsymbol{x} \in X \mid \exists \text{ some } \varepsilon > 0 \text{ such that } N_\varepsilon(\boldsymbol{x}) \subseteq X\}$$

[pictures: $\mathbb{R}$, $\mathbb{R}^2$]

SOLUTIONS

A **solution** to the problem $\max\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in X\}$ or **maximizer** of f on X is a point $\boldsymbol{x}^*$ s.t.

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in X.$$

A **solution** to the problem $\min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in X\}$ or **minimizer** of f on X is a point $\boldsymbol{x}^*$ s.t.

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in X.$$

LOCAL VS. GLOBAL SOLUTIONS

$\boldsymbol{x}^* \in X$ is a **global maximizer** of f in X iff $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in X$.

$\boldsymbol{x}^* \in X$ is a **global minimizer** of f in X iff $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in X$.

$\boldsymbol{x}^* \in X$ is a **local maximizer** of f in X iff $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in N_\varepsilon(\boldsymbol{x}^*) \cap X$.

$\boldsymbol{x}^* \in X$ is a **local minimizer** of f in X iff $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in N_\varepsilon(\boldsymbol{x}^*) \cap X$.

[pictures]

Most of the definitions or results for maximization problems have an exact analog for minimization problems. From now on, the minimization analog will be omitted.

A solution $\boldsymbol{x}^* \in X$ is **interior** iff there is an $\varepsilon > 0$ such that $N_\varepsilon(\boldsymbol{x}^*) \subseteq X$ and $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ $\forall \boldsymbol{x} \in N_\varepsilon(\boldsymbol{x}^*)$.

[picture]

SET OF SOLUTIONS

The **set of solutions** to a max problem is denoted

$$\text{argmax}\,(f(\boldsymbol{x}) \mid \boldsymbol{x} \in X) \;=\; \{\boldsymbol{x} \in X \mid f(\boldsymbol{x}) \geq f(\boldsymbol{x}') \;\forall\; \boldsymbol{x}' \in X\}$$

Consider the following example:

Let $X = [-1, 1]$, and $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$. [picture]

Maximizing f on C has two solutions, $x = -1$ and $x = 1$.

As we can see, the set $\arg\max(f(\boldsymbol{x}) \mid \boldsymbol{x} \in X)$ can have more than one element.

Example 1

Let $X = [0,1]$, and $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$.

[picture]

The problem of maximizing f on X has exactly one solution, the point $x = 1$.

The set $\arg\max(f(\boldsymbol{x}) \mid \boldsymbol{x} \in X)$ can be a singleton.

EXAMPLE 2

Let $X = \mathbb{R}_+$, and $f : \mathbb{R}_+ \to \mathbb{R}$ be $f(x) = x^2$.

[picture]

The problem of maximizing f on X has no solution.

The set $\arg\max(f(x) \mid x \in X) = \varnothing$.

EXISTENCE OF A SOLUTION

A sufficient condition for existence of a solution is given by the Weierstrass theorem.

THEOREM: (**Weierstrass**) If f is continuous and X is closed and bounded (hence compact) and nonempty, then there exist a global maximum and a global minimum.

However, Weierstrass is sufficient but not necessary for existence of a solution.

Example: $X = (0, 2]$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ be $f(x) = x^2$

[picture]

OBJECTIVES OF OPTIMIZATION THEORY

I. To identify a set of conditions on f and X that guarantee the **existence** of solutions to optimization problems. [Weierstrass theorem]

II. To obtain a characterization of the set of solutions:

A. Necessary conditions [picture] [differentiability]
B. Sufficient conditions [picture] [convexity]
C. Conditions that guarantee uniqueness of a solution
D. A theory of parametric variation: [envelope theorem]

Sometimes optimization problems are presented in parametric form: f and or X depend on parameters $\boldsymbol{\theta} \in \Theta$, where Θ (set of feasible parameter values). We write: $\max\{f(\boldsymbol{x}, \boldsymbol{\theta}) \mid \boldsymbol{x} \in X(\boldsymbol{\theta})\}$

MOTIVATION (1)

Consumer's Utility Maximization Problem

$$\begin{aligned} \max \ & u(\boldsymbol{x}) \\ \text{s.t.} \ & \boldsymbol{p} \cdot \boldsymbol{x} \leq m \\ & \boldsymbol{x} \geq \mathbf{0} \\ & \boldsymbol{p} \in \mathbb{R}^l_{++} \\ & \boldsymbol{x} \in \mathbb{R}^l \end{aligned}$$

Does a solution exist?

The constraint set $X = B(\boldsymbol{p}, m) = \{\boldsymbol{x} \in \mathbb{R}^l : p \cdot \boldsymbol{x} \leq m\}$ is compact. [picture]

So if u is continuous, then a solution exists (by Weierstrass theorem).

Motivation (2)

Firm's Cost Minimization Problem

$$\begin{gathered} \min \; \boldsymbol{p} \cdot \boldsymbol{x} \\ \text{s.t. } f(\boldsymbol{x}) \geq y \\ \boldsymbol{x} \geq 0 \\ \boldsymbol{p} \in \mathbb{R}^l_{++}, \; y \in \mathbb{R}_{++} \\ \boldsymbol{x} \in \mathbb{R}^l \end{gathered}$$

Does a solution exist?

The constraint set is *not* compact. [picture]

If the constraint set is not empty, $\exists\, \boldsymbol{x}^*$ s.t. $f(\boldsymbol{x}^*) \geq y$.

"Compactify" the constraint set: $\{\boldsymbol{x} : \boldsymbol{p} \cdot \boldsymbol{x} \leq \boldsymbol{p} \cdot \boldsymbol{x}^* \text{ and } f(\boldsymbol{x}) \geq y\}$.
Since the objective function is continuous, there exists a solution.

USEFUL THEOREMS (1)

THEOREM: Let $-f$ denote the function whose value at any $\boldsymbol{x}$ is $-f(\boldsymbol{x})$. Then $\boldsymbol{x}$ is a maximum of f on X iff it is a minimum of $-f$ on X; and $\boldsymbol{z}$ is a minimum of f on X iff $\boldsymbol{z}$ is a maximum of $-f$ on X.

PROOF: HW. ■

USEFUL THEOREMS (2)

THEOREM: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function, that is, a function such that

$$y > y' \Rightarrow \varphi(y) > \varphi(y')$$

Then $\boldsymbol{x}$ is a maximum of f on X iff $\boldsymbol{x}$ is a maximum of the composition of φ and f on X; and $\boldsymbol{z}$ is a minimum of f on X iff $\boldsymbol{z}$ is a minimum of $\varphi \circ f$ on X.

REMARK: it suffices that φ be a strictly increasing function on just the set $f(X)$.That is, that φ only satisfies $\varphi(y) > \varphi(y')$ for all y, $y' \in f(X)$ with $y > y'$.

PROOF: HW. ■

Constrained Optimization

Consider the problem (P):

$$\max f_0(\boldsymbol{x})$$
$$\text{s.t.}$$
$$\boldsymbol{x} \in S;$$
$$\mathrm{f}_1(\boldsymbol{x}) \geq 0, \; \dots \; f_m(\boldsymbol{x}) \geq 0$$

Where:

- $S \subset \mathbb{R}^N$ is convex
- $f_i : S \to \mathbb{R}$ for $i = 0, 1, \dots m$ are differentiable and concave

EXAMPLE

Consumer's Utility Maximization Problem

$$\begin{aligned} \max \;& u(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{x} \in \mathbb{R}^l_+ \\ & m - \boldsymbol{p} \cdot \boldsymbol{x} \geq 0 \\ \text{where } & \boldsymbol{p} \in \mathbb{R}^l_{++} \end{aligned}$$

SLATER CONDITION

Slater Condition: $\exists\, \hat{\boldsymbol{x}} \in \text{int}(S)$ s.t. $f_i(\hat{\boldsymbol{x}}) > 0$ for $i = 1, ...m$.

EXAMPLE:

$f_0(x) = x,\ f_1(x) = -x^2$

Let $S = \mathbb{R}$

So the maximization problem is:

$$\begin{aligned} \max_x \quad & x \\ \text{s.t.} \quad & x \in \mathbb{R} \\ & -x^2 \geq 0 \end{aligned}$$

There is no $x \in \mathbb{R}$ s.t. $-x^2 > 0$ so Slater is violated.

THE LAGRANGEAN

Associated with (P), we can define a function $L : S \times \mathbb{R}^m_+ \to \mathbb{R}$ given by:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) \; = \; f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x})$$

Notice that, for a given value of $\boldsymbol{\lambda}$, $L(\boldsymbol{x}, \cdot)$ is concave in $\boldsymbol{x}$, and for a given value of $\boldsymbol{x}$, $L(\cdot, \boldsymbol{\lambda})$ is convex in $\boldsymbol{\lambda}$.

A function with these properties is also called **saddle function**.

SADDLE POINT

DEFINITION: a point $(\boldsymbol{x}^*, \lambda^*)$ is a **saddle point** of $L(\boldsymbol{x}, \boldsymbol{\lambda})$ if

$$L(\boldsymbol{x}, \boldsymbol{\lambda}^*) \leq L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \leq L(\boldsymbol{x}^*, \boldsymbol{\lambda})$$

for all $(\boldsymbol{x}, \boldsymbol{\lambda}) \in S \times \mathbb{R}^m_+$

SADDLE POINT: CHARACTERIZATION

If $\boldsymbol{x}^* \in \text{int}(S)$, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point iff

1. $Df_0(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* Df_i(\boldsymbol{x}^*) = 0$
2. $f_i(\boldsymbol{x}^*) \geq 0, \ \lambda_i^* \geq 0$ for $i = 1, ...m$
3. $\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*) = 0$

Given condition 2, condition 3 can be replaced by

3′. $\lambda_i^* f_i(\boldsymbol{x}^*) = 0$

KUHN TUCKER UNDER CONCAVITY

THEOREM: **(Kuhn-Tucker I)** Assume $f_0, f_1, ..., f_M$ are concave, continuous functions from a convex set $S \subseteq \mathbb{R}^N$ into $\mathbb{R}$. Let the problem (P) and $L(\boldsymbol{x}, \boldsymbol{\lambda})$ be as described above. Then:

(i) If $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \in S \times \mathbb{R}^M$ is a saddle point of $L(\boldsymbol{x}, \boldsymbol{\lambda})$, then $\boldsymbol{x}^*$ solves (P).

(ii) Assume that the Slater condition holds. Then if $\boldsymbol{x} \in S$ is a solution to (P), $\exists \boldsymbol{\lambda}^* \in \mathbb{R}^M$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of $L(\boldsymbol{x}, \boldsymbol{\lambda})$.

This version of Kuhn Tucker doesn't require differentiability.

EFFECTIVE CONSTRAINTS

DEFINITION: an inequality constraint is **effective** or **binding** at a certain point $\boldsymbol{x}^*$ if $f_i(\boldsymbol{x}^*) = 0$ i.e. if the constraint holds with equality at $\boldsymbol{x}^*$.

KUHN TUCKER WITHOUT CONCAVITY

THEOREM: **(Kuhn Tucker II)** Let $f_0 : S \to \mathbb{R}$ be a C^1 function on a certain open set $S \subseteq \mathbb{R}^N$, and let $f_i : \mathbb{R}^N \to \mathbb{R},\ i = 1, ..., M$ be C^1 functions. Suppose that $\boldsymbol{x}^*$ is a local maximum of f_0 on the set $D = S \cap \{\boldsymbol{x} \in \mathbb{R}^N | f_i(\boldsymbol{x}) \geq 0,\ i = 1, ...M\}$.

Let $E \subseteq \{1, ..., M\}$ denote the set of effective constraints at $\boldsymbol{x}^*$. Suppose that the derivatives $\{Df_i(\boldsymbol{x}^*)|i \in E\}$ form an independent set of vectors. Then there exist $\lambda_i^* \in \mathbb{R},\ i = 1, ..., M$ s.t.

(i) $\lambda_i^* \geq 0,\ i = 1, ..., M;$

(ii) $\lambda_i^* f_i(\boldsymbol{x}^*) = 0,\ i = 1, ..., M;$

(iii) $Df_0(\boldsymbol{x}^*) + \sum_{i=1}^{M} \lambda_i f_i(\boldsymbol{x}^*) = 0$

NOTE: these conditions are only necessary.

Another Approach

State problems in "standard form":

MAX PROBLEM

$$\begin{aligned} &\max \ f(\boldsymbol{x}) \\ \text{s.t.} \ &g_1(\boldsymbol{x}) \le 0 \\ &\vdots \\ &g_K(\boldsymbol{x}) \le 0 \end{aligned}$$

MIN PROBLEM

$$\begin{aligned} &\min \ f(\boldsymbol{x}) \\ \text{s.t.} \ &g_1(\boldsymbol{x}) \ge 0 \\ &\vdots \\ &g_K(\boldsymbol{x}) \ge 0 \end{aligned}$$

Example

(i) Consumer:

$$\max \ u(\boldsymbol{x})$$
$$\text{s.t.} \ \ \boldsymbol{p} \cdot \boldsymbol{x} - m \leq 0$$
$$-x_1 \leq 0$$
$$\vdots$$
$$-x_N \leq 0$$

(ii) Firm:

$$\min \ \boldsymbol{p} \cdot \boldsymbol{x}$$
$$\text{s.t.} \ \ f(\boldsymbol{x}) - y \geq 0$$
$$x_1 \geq 0$$
$$\vdots$$
$$x_N \geq 0$$

Constraint qualification

Recall we defined the set E which contains only the indices of the binding (effective) constraints.

Definition: **constraint qualification (CQ)** holds at $\boldsymbol{x}^*$ iff $\{\nabla g_i(\boldsymbol{x}^*) : i \in E\}$ is independent.

KUHN TUCKER II RESTATED

THEOREM **(KT)**: Consider a MAX problem in the standard form. Let $f,\ g_k$ be C^1. Let $\boldsymbol{x}^*$ be a local maximum. Suppose that CQ holds at $\boldsymbol{x}^*$. Then $\exists\ \lambda_k \geq 0$ s.t.

$$(1) \quad \nabla f(\boldsymbol{x}^*) = \sum_{k=1}^{K} \lambda_k^* \nabla g_k(\boldsymbol{x}^*)$$

$$(2) \quad \lambda_k^* g_k(\boldsymbol{x}^*) = 0 \quad \forall k \quad \text{(complementary slackness)}$$

NOTE: (2) says that if $g_k(\boldsymbol{x}^*) < 0$ (kth constraint not binding), then $\lambda_k = 0$. We can rewrite (1):

$$\nabla f(\boldsymbol{x}^*) = \sum_{k \in E} \lambda_k^* \nabla g_k(\boldsymbol{x}^*)$$

EXAMPLE

Let $f(x_1, x_2) = x_2$. Let the constraints be

$g_1 = (x_1 - 1)^2 + x_2^2 \leq 1$ [a disk centered at $(1, 0)$]

$g_2 = (x_1 + 1)^2 + x_2^2 \leq 1$ [a disk centered at $(-1, 0)$]

The only feasible point is $(0, 0)$, so it is the solution.

$$\nabla f(\hat{x}^*) = (0, 1)$$

$$\nabla g_1(x^*) = (-2, 0)$$

$$\nabla g_2(x^*) = (2, 0)$$

EXAMPLE (2)

There's no way of writing $\nabla f(x^*)$ as a linear combination of $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$!

Condition KT(1) fails.

What is wrong with this example?

Constraint qualification (CQ) fails at $(0, 0)$.

But CQ isn't necessary.

There are other restrictions on the MAX problem sufficient to guarantee KT conditions hold under other assumptions (see previous section, Kuhn Tucker under concavity)

Binding vs. active constraints

Example:

max $f(x) = -x^2$ s.t. $g(x) = -x \leq 0$ [picture]

The solution is $x^* = 0$. The constraint is binding, but $\lambda = 0$ because $\nabla f(x^*) = 0$. Relaxing the constraint does not change the solution.

Call constraint k **active** if $\lambda_k > 0$.

Then g in the previous example is binding, but not active.

If a constraint is active, then it is binding (by condition KT (2)). Most of the times, binding constraints will be active, but not always.

SLACK CONSTRAINTS CAN AFFECT THE GLOBAL SOLUTIONS

EXAMPLE:

$$\max \ f(x) = -x^2 + x^4 \quad \text{s.t.} \ g_1(x) = -x - 1 \leq 0,$$
$$g_2(x) = x - 1 \leq 0$$

[picture: "W"]

We have three constrained maxima at -1, 0 and 1.

At $x^* = 0$, none of the constraints is binding: $g_1(x^*) < 0$ and $g_2(x^*) < 0$.

However, if, say, we relax g_2 to $x - 3 \leq 0$, then there would be a unique maximum at $x^* = 3$.

KT is a result about local rather than global maximization. Even if we relax the constraint, $x^* = 0$ remains a local maximum.

INTUITION: ONE BINDING CONSTRAINT

Mountain example.

By (1), $\nabla f(\boldsymbol{x}^*) = \lambda \nabla g(\boldsymbol{x}^*)$

- $g(\boldsymbol{x}^*) = 0$: g cannot be increased.
- $\nabla g(\boldsymbol{x}^*) \perp$ level set $g(\boldsymbol{x})^*$ (the fence).
  (we can only move along the level set $g(\boldsymbol{x}^*)$).
- $\nabla f(\boldsymbol{x}^*)$ is also $\perp$ level set $g(\boldsymbol{x})^*$
  (any movement that we're allowed to make does not increase f).

[pictures]

Example 1

$\max f(x) = \sqrt{x}$ s.t. $x \geq 0$, x≤ 1

[picture]

$$\lambda_1(\overbrace{-x}^{g_1} \leq 0)$$
$$\lambda_2(\underbrace{x-1}_{g_2} \leq 0)$$

At $x^* = 1$: $\nabla f(x) = Df(x) = \frac{1}{2\sqrt{x}}$

$\nabla f(x^*) = \frac{1}{2}$

$\lambda_1 = 0$ by KT (2).

λ_2 solves $\nabla f(x^*) = \lambda_2^* \nabla g_2(x^*)$

$\frac{1}{2} = \lambda_2(1) \Rightarrow \lambda_2 = \frac{1}{2}$

[picture of the gradients]

Example 2

$$\max f(x) = e^{-x} \quad \text{s.t.} \quad x \geq 0 \quad (-x \leq 0) \qquad \text{[picture]}$$

At $x^* = 0$:
$$\nabla f(x) = -e^{-x}$$
$$\nabla f(x^*) = -1$$
$$\nabla g(x^*) = -1$$

By (2), $\lambda = 1$

[picture of gradients]

Example 3

For $N = 2$, suppose (1) doesn't hold ($\nabla f(\boldsymbol{x}^*)$ and $\nabla g(\boldsymbol{x}^*)$are not collinear) Then it's feasible to move and increase f simultaneously.

[picture]

INTUITION: TWO BINDING CONSTRAINTS

[picture]

Mountain example.

By (1), $\nabla f(\boldsymbol{x}^*) = \lambda_1^* \nabla g_1(\boldsymbol{x}^*) \; + \; \lambda_2^* \nabla g_2(\boldsymbol{x}^*)$.

The gradient of the objective function lies in the cone spanned by the gradients of each binding constraint.

[picture]

PROOF OF KUHN TUCKER

Define $W = \{\boldsymbol{x} \in \mathbb{R}^N : \exists \lambda_k \text{ s.t. } \boldsymbol{x} = \sum \lambda_k \nabla g_k(\boldsymbol{x}^*)_{\boldsymbol{x} \in E}\}$

WTS: if $\boldsymbol{x}^*$ is a local max, then $\nabla f(\boldsymbol{x}^*) \in W$.

By contraposition: if $\nabla f(\boldsymbol{x}^*) \notin W$, then $\boldsymbol{x}^*$ can't be a local max.

Suppose $\nabla f(\boldsymbol{x}^*) \notin W$. Since W is convex and closed, and $\nabla f(\boldsymbol{x}^*)$ is compact and convex (a point), by the (strict) Separating Hyperplane Theorem, $\exists\, \boldsymbol{v} \neq 0 \in \mathbb{R}^N$ and $c \in \mathbb{R}$ such that

$$\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{v} > c > \boldsymbol{w} \cdot \boldsymbol{v} \qquad \forall\, w \in W$$

Proof of Kuhn Tucker (continued – 1)

(i) Since $\mathbf{0} \in W$ (set all $\lambda_k = 0$), $c > 0$, so $\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{v} > 0$.

Given $\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{v} > 0$, $D_{\boldsymbol{v}} f(\boldsymbol{x}^*) > 0$

(there is a movement in direction $\boldsymbol{v}$ that increases f)

(ii) Since $\lambda_k \nabla g_k(\boldsymbol{x}^*) \in W$, then $c > \lambda_k \nabla g_k(\boldsymbol{x}^*) \cdot \boldsymbol{v}$

For any $\lambda_k > 0$, $c/\lambda_k > \nabla g_k(\boldsymbol{x}^*) \cdot \boldsymbol{v}$

Taking limits as $\lambda \to \infty$, $0 \geq \nabla g_k(\boldsymbol{x}^*) \cdot \boldsymbol{v} \quad (\forall k \in E)$

The movement in direction $\boldsymbol{v}$ is feasible.

Proof of Kuhn Tucker (continued – 2)

Define $J = \{k \in E : \nabla g_k(\boldsymbol{x}^*) \cdot \boldsymbol{v} = 0\}$

If $J = \emptyset$, then $\nabla g_k(\boldsymbol{x}^*) \cdot \boldsymbol{v} < 0 \quad \forall k \in E$

$\forall\, \alpha$ sufficiently small, $g_k(\boldsymbol{x}^* + \alpha \boldsymbol{v}) < g_k(\boldsymbol{x}^*) = 0$ (g_k no longer binds);

for $k \notin E$, $g_k(\boldsymbol{x}^* + \alpha \boldsymbol{v}) < 0$, so by continuity of g, $g_k(\boldsymbol{x}^* + \alpha \boldsymbol{v}) < 0 \;\; \forall \alpha$ sufficiently small.

So $\forall k$, $g_k(\boldsymbol{x}^* + \alpha \boldsymbol{v}) < 0 \quad \forall \alpha$ sufficiently small, so the movement is feasible.

Also, given that $\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{v} > 0$, then $\nabla f(\boldsymbol{x}^* + \alpha \boldsymbol{v}) > \nabla f(\boldsymbol{x}^*)$ for α sufficiently small, so $\boldsymbol{x}^*$ cannot be a local maximum.

PROOF OF KUHN TUCKER (CONTINUED – 3)

If $J \neq \emptyset$, then $\exists$ at least one $k \in E$ for which $\nabla g_k(\boldsymbol{x}^*)v = 0$, so the point $\boldsymbol{x}^* + \alpha \boldsymbol{v}$ might not be feasible.

Use Implicit Function Theorem to argue that $\exists$ points $\boldsymbol{x}$ that are (a) feasible, (b) arbitrarily close to $\boldsymbol{x}^*$, and (c) the movement from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ is arbitrarily close to $\boldsymbol{v}$.

Let $K = |J|$. Let $S = \{\nabla g(\boldsymbol{x}^*) : k \in J\}$. By CQ, S is independent.

By CQ and Implicit Fcn Thm, equations $g_k(\boldsymbol{x}^*) = 0 \;\; \forall k \in J$ implicitly define a C^1 function $\psi : \mathbb{R}^{(N-K)} \to \mathbb{R}^N$ s.t. $\forall k \in J$ and $\boldsymbol{z} \in \mathbb{R}^{(N-K)}$, $g_k(\psi(\boldsymbol{z})) = 0$, $D\psi(\boldsymbol{z})$ has full rank, and $\psi(\boldsymbol{0}) = \boldsymbol{x}^*$

That is, ψ gives a $N-K$-dimensional surface consisting of all points near $\boldsymbol{x}^*$ for which the constraints in J hold with equality.

PROOF OF KUHN TUCKER (CONTINUED – 4)

Given that $g_k(\psi(\boldsymbol{z})) = 0 \;\; \forall \boldsymbol{z} \in \mathbb{R}^{N-K}$ and all $k \in J$, by the Chain Rule, $Dg_k(\boldsymbol{x}^*)D\psi(\boldsymbol{0})\boldsymbol{z} = \boldsymbol{0} \;\; \forall \boldsymbol{z} \in \mathbb{R}^{N-K}$ and all $k \in J$.

Let $A = [D_k]_{k \in J}$, $(K \times N)$. By CQ, A has full rank (K), so its null space has dimension $N - K$.

Since $D\psi$ has full rank $(N - K)$, it maps onto the null space of A.

Since $\boldsymbol{v}$ lies in the null space of A, $\exists \boldsymbol{z}_v \in \mathbb{R}^{(N-K)}$ s.t. $\boldsymbol{v} = D\psi(\boldsymbol{0})\boldsymbol{z}_v$.

For sufficiently small $\alpha > 0$, $\psi(\alpha \boldsymbol{z}_v)$ gives such point $\boldsymbol{x}$ that fulfills (a), (b) and (c).

PROOF OF KUHN TUCKER (CONTINUED – 5)

We already know that $\forall\, k \in J,\ g_k(\psi(\alpha \boldsymbol{z}_v) = 0\ \ \forall \alpha > 0.$

For $k \in E \backslash J$, by Chain Rule, $Dg_k(\boldsymbol{x}^*)D\psi(\mathbf{0})\boldsymbol{z}_v = Dg_k(\boldsymbol{x}^*)\boldsymbol{v} < 0$, which implies that $g_k(\psi(\alpha \boldsymbol{z}_v)) < g_k(\boldsymbol{x}^*) = 0$ for α sufficiently small.

For $k \notin E$, $g_k(\boldsymbol{x}^*) < 0$. By continuity, $g_k(\psi(\alpha \boldsymbol{z}_v)) < 0$ for α sufficiently small.

Thus, $\boldsymbol{x} = \psi(\alpha z)$ is feasible.

By Chain Rule, $Df(\boldsymbol{x}^*)D\psi(\mathbf{0})\boldsymbol{z}_v = Df(\boldsymbol{x}^*)\boldsymbol{v}$. Since $\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{v} > 0$, $Df(\boldsymbol{x}^*)\boldsymbol{v} > 0$

Hence, for α sufficiently small, $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$. ∎

USING THE KT CONDITIONS

Bad news! There is no easy way of finding points $\boldsymbol{x}^*$ and multipliers λ that satisfy the KT conditions.

Cookbook?

Try solving the unconstrained problem first. If the solution satisfies the constraints, you're done.

If it doesn't, make an educated guess of which of the constraints might be binding.

Example 1

$$\begin{aligned} \max \quad & f(x) = x_1^{1/2} x_2^{1/3} x_3^{1/6} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 \leq 9 \\ & x \geq 0 \end{aligned}$$

In the standard form:

$$\begin{aligned} \max \quad & f(x) = x_1^{1/2} x_2^{1/3} x_3^{1/6} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -x \leq 0 \end{aligned}$$

EXAMPLE 1

i. None of the non-negativity constraints will bind.

How do we know?

If any of the $x_n = 0$, then $f(x) = 0$.
Take any feasible point, for instance, $(\frac{1}{4}, \frac{1}{8}, \frac{1}{3})$
plug it into the constraint: $3 - 9 = -6 \leq 0$,
which yields $f(x) = (\frac{1}{2})^2(\frac{1}{3})^{1/6} > 0$.

ii. The constraint $4x_1 + 8x_2 + 3x_3 - 9 = 0$ is binding.

How do we know?

$Df(x) >> 0$.
From KT(2), we know that $\lambda_2 = \lambda_3 = \lambda_4 = 0$.
Use KT(1) to find λ_1.

Example 1

iii. Want to avoid messy calculations?

Use the second "useful theorem": apply a strictly increasing transformation to $f(\boldsymbol{x})$, and solve the max problem.

Let $\hat{f}(\boldsymbol{x}) = Ln(f(\boldsymbol{x}))$. This transformation is strictly increasing: $D(Ln(y)) = \frac{1}{y} > 0$ (we already know that $f(\boldsymbol{x}) > 0$)

So solving the problem:

$$\begin{aligned} \max \quad & \hat{f}(\boldsymbol{x}) = \tfrac{1}{2}\ln x_1 + \tfrac{1}{3}\ln x_2 + \tfrac{1}{6}\ln x_3 \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -\boldsymbol{x} \leq 0 \end{aligned}$$

Yields the same solution as the original problem.

Example 1

Use KT(1): $\begin{bmatrix} \frac{1}{2x_1^*} \\ \frac{1}{3x_2^*} \\ \frac{1}{6x_3^*} \end{bmatrix} = \lambda_1 \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix}$ or $\begin{bmatrix} \frac{1}{2\lambda_1} \\ \frac{1}{3\lambda_1} \\ \frac{1}{6\lambda_1} \end{bmatrix} = \begin{bmatrix} 4x_1^* \\ 8x_2^* \\ 3x_3^* \end{bmatrix}$

Substituting in the constraint, we get $\frac{1}{2\lambda_1} + \frac{1}{3\lambda_1} + \frac{1}{6\lambda_1} = 9, \quad \lambda_1 = \frac{1}{9}$

Substituting back, $\boldsymbol{x}^* = \begin{bmatrix} \frac{9}{8} \\ \frac{3}{8} \\ \frac{1}{2} \end{bmatrix}$

However, I haven't shown yet that x^* is a solution. So far, we only know that it satisfies the KT necessary conditions. To show that x^* is a solution, I need sufficient conditions.

Example 2

$$\begin{aligned} \max\ & f(x) = \sqrt{x_1+1} + \sqrt{x_2+1} + \sqrt{x_3+1} \\ \text{s.t. } & 4x_1 + 8x_2 + 3x_3 \leq 9 \\ & x \geq 0 \end{aligned}$$

In standard form:

$$\begin{aligned} \max\ & f(x) = \sqrt{x_1+1} + \sqrt{x_2+1} + \sqrt{x_3+1} \\ \text{s.t. } & g_1 = 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -x \leq 0 \end{aligned}$$

Now it's no longer obvious that the solution has $x >> 0$.

What do we know now?

(i) $Df(x) >> 0$, so the first constraint will bind:
$4x_1 + 8x_2 + 3x_3 - 9 = 0$.

EXAMPLE 2

Now start guessing!

Guess 1: $x >> 0$?

If only the first constraint binds, then by KT(2), $\lambda_2 = \lambda_3 = \lambda_4 = 0$

And by KT(1), $\begin{bmatrix} \frac{1}{2\sqrt{x_1^*+1}} \\ \frac{1}{\sqrt{x_2^*+1}} \\ \frac{3}{2\sqrt{x_3^*+1}} \end{bmatrix} = \lambda_1 \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix}$

After some calculations we get $\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} \\ -\frac{3}{5} \\ \frac{27}{5} \end{bmatrix}$ and $\lambda_1 = \frac{\sqrt{10}}{16}$

But this point is not in the feasible set!

Example 2

This bad guess can give us a clue of what the solution looks like.

Guess 2: $x_1 = x_2 = 0$

Then, since the first constraint binds, $x_3 = 3$, so $x^* = (0, 0, 3)$

$$\nabla f(x^*) = \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{3}{4} \end{bmatrix} \; \nabla g_1(x^*) = \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix}, \quad \nabla g_2(x^*) = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \quad \nabla g_3(x^*) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

Write $\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) + \lambda_3 \nabla g_3(x^*)$ in matrix form:

$$\begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 4 & -1 & 0 \\ 8 & 0 & -1 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{bmatrix} > 0$$

Set $\lambda_4 = 0$, and both KT(1) and KT(2) hold.

EXAMPLE 2

What would have happened if I had guessed that $x_2 = x_3 = 0$?

$\Rightarrow x^* = (\frac{9}{4}, 0, 0)$, which is feasible, so we would expect KT to fail.

By KT(2), $\lambda_2 = 0$

By KT(1), $\begin{bmatrix} \lambda_1 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \approx \begin{bmatrix} 0.1 \\ -0.4 \\ -1.4 \end{bmatrix}$

So $\lambda_k \geq 0$ is violated.

[The sign of KT multipliers is important].

KT can catch bad guesses.